



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2019

---

## **Confounders in epidemiological associations**

Papageorgiou, Spyridon N

**Abstract:** The following description of a retrospective non-randomized study is given within a publication in a scientific journal: “A total of 54 consecutive patients (24 male / 28 female; mean age: 14.1 years) treated with sequential thermoplastic aligners during the last 12 months were identified from the archive of a private practice. They were compared to 52 consecutive patients treated with conventional fixed appliances during the same period, who were matched for age, sex, and case severity with the patients in the aligner group. The total duration of treatment in months was extracted from the patient files by a third party not involved in any way with their treatment. Initially, descriptive statistics were calculated for all patient characteristics and for the study’s primary outcome (treatment duration), consisting of means and Standard Deviations (SDs). Student’s t-tests for independent samples and chi-square tests were performed on patient age, sex, and case severity to confirm that the two groups were matched. Finally, a t-test for independent samples was performed on treatment duration to assess any differences between the aligner and the fixed appliance group at the 5% level.” The authors of the study give the following table (Table 1) in their Results section and conclude that (i) the two groups were adequately matched, since no statistically significant difference was found for any baseline difference and (ii) aligners and braces are equally efficient, since no statistically significant difference was found for treatment duration.

DOI: <https://doi.org/10.1177/1465312519869934>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-186075>

Journal Article

Accepted Version

Originally published at:

Papageorgiou, Spyridon N (2019). Confounders in epidemiological associations. *Journal of Orthodontics*, 46(3):267-269.

DOI: <https://doi.org/10.1177/1465312519869934>

## **Confounders in epidemiological associations**

Spyridon N. Papageorgiou

Clinic of Orthodontics and Pediatric Dentistry, Center of Dental Medicine, University of Zurich, Zurich, Switzerland

ORCID: Spyridon N. Papageorgiou <http://orcid.org/0000-0003-1968-3326>

CONTACT: Spyridon N. Papageorgiou Clinic of Orthodontics and Pediatric Dentistry, Center of Dental Medicine, University of Zurich, Plattenstrasse 11, Zurich CH 8032, Switzerland; [snpapage@gmail.com](mailto:snpapage@gmail.com).

Words in text: 1350

Disclosure statement

No potential conflict of interest was reported by the author.

## Theoretical scenario

The following description of a retrospective non-randomized study is given within a publication in a scientific journal:

*“A total of 54 consecutive patients (24 male / 28 female; mean age: 14.1 years) treated with sequential thermoplastic aligners during the last 12 months were identified from the archive of a private practice. They were compared to 52 consecutive patients treated with conventional fixed appliances during the same period, who were matched for age, sex, and case severity with the patients in the aligner group. The total duration of treatment in months was extracted from the patient files by a third party not involved in any way with their treatment. Initially, descriptive statistics were calculated for all patient characteristics and for the study’s primary outcome (treatment duration), consisting of means and Standard Deviations (SDs). Student’s t-tests for independent samples and chi-square tests were performed on patient age, sex, and case severity to confirm that the two groups were matched. Finally, a t-test for independent samples was performed on treatment duration to assess any differences between the aligner and the fixed appliance group at the 5% level.”*

The authors of the study give the following table (Table 1) in their Results section and conclude that (i) the two groups were adequately matched, since no statistically significant difference was found for any baseline difference and (ii) aligners and braces are equally efficient, since no statistically significant difference was found for treatment duration.

### Which of the following statements are correct, if any?

- (a) The variables of patient age, sex, and case severity all act as confounders to the relationship between appliance choice (aligners versus fixed appliances) and treatment duration.
- (b) The two groups were adequately matched for patient age, sex, and case severity, as all p values for baseline differences were non-significant.
- (c) We can be confident that no difference in treatment duration exists between aligners and fixed appliances, since the two groups were matched at baseline and this minimizes the impact of confounders.
- (d) The two groups were adequately matched for all possible confounders that could have affected the study’s outcome.

## Discussion

In order to judge if the abovementioned statements are true or not, we need to define what a confounder is. According to a specific definition, as confounder can be considered any covariate that is either a cause for an outcome or a surrogate (proxy) for an outcome (Greenland and Morgenstern, 2001). Or to put it another way, as confounder can be considered any variable for which adjustment is helpful in reducing bias in effect estimation. To better illustrate how confounders work, consider the following example: suppose we are given the city statistics covering a 5-month summer period and we work out statistically that the number of ice creams sold per day correlate significantly with the number of swimming pool deaths per day (Westfall and Yarkoni, 2016). Would we then reach the conclusion that ice cream is a risk factor for your drowning? No, because we are missing out from the picture the intermediary role that recorded daily temperature plays. On hot days people tend to both eat more ice creams and also tend to go more often swimming in pools (and subsequently have higher risk of drowning) than on cold days. Therefore, when looking at the effect of ice cream sales/day on the number of drownings/day, the recorded daily temperature acts as a confounder – that is to say, it introduces bias in estimating an effect.

Based on this knowledge it becomes obvious that confounders are specific to the relationship and the outcome we are interested in. Therefore, coming back to our theoretical study of aligners versus fixed appliances, as confounder we would consider any covariate that might influence the effect of appliance choice on duration of treatment. One way is to look at the effect of patient demographics on treatment duration through simple linear regression that gives for each association an unstandardized regression coefficient ( $\beta$ ), its 95% Confidence Interval (CI), and a p value (even though this is a very crude method and much better ways exist to do this). It seems that patient age ( $\beta=0.11$ ; 95% CI=-0.43 to 0.66; P value=0.69) is not a statistically significant predictor of treatment duration at the 5% level. Patient sex ( $\beta=-1.51$ ; 95% CI=-3.24 to 0.21; P value=0.09) might exert some influence on treatment duration since a considerable difference of -1.51 month is seen between male-female patients, but it does not formally cross the 5% level. On the other side, many studies in the literature report that the more 'difficult' a case is at the start, the longer it takes to treat the case, which makes sense. One widely-used method to assess baseline case severity (especially in the United Kingdom) is through the Peer Assessment Rating (PAR) index

Richmond et al., 1992), which the authors of the study also measured. Indeed, baseline case severity (assessed with the PAR index) was significantly associated with treatment duration ( $\beta=0.34$ , 95% CI=0.30 to 0.39; P value<0.001; Fig 1). The interpretation of this relationship is that an average increase of 0.34 month in treatment duration can be expected for each additional PAR point at baseline. Therefore, using the rather crude method of univariable linear regression, we can conclude that baseline severity might act as a confounder on the relationship between appliance choice on treatment duration, while patient age and sex didn't. Therefore, statement (a) is wrong.

The authors of the study concluded that the two groups were matched for all confounders, since all P values for baseline differences were non-significant at the 5% level. This would probably mean that if we would control for any of these confounders, we would still reach similar conclusions to the authors about the study's outcome.

One way to take into account any possible confounders acting on a relationship is to 'control' for them statistically – provided of course that we know which variables act as confounders and that we have reliably measured these variables. This can be seen in Table 2, where several regression models with appliance type as independent variable and treatment outcome as dependent variable are given. The 'simple' model 1 includes only these two variables, whereas patient age, sex, and baseline PAR scores are 'controlled' for in the 'adjusted' models 2-4, respectively. One can see that in the simple model 1 no significant difference in treatment duration is seen between aligners and fixed appliances, which is the authors' conclusion. Similar observations are made for models 2 and 3 – i.e. when controlling for the effect of patient age or sex. On the other side, model 4 provides a different take on the data. When controlling for the effect of baseline severity within the assessed sample, treatment with braces is significantly shorter by -1.3 month compared to aligners (95% CI=-2.3 to -0.3 months; P value=0.01). The fact therefore that the p value for baseline differences between groups was not smaller than 0.05 doesn't necessarily mean that the groups were fully matched. Consequently, the authors initial conclusion of similar treatment times for aligners and braces seems to have been biased by the groups' baseline severity. Therefore, both statement (b) and statement (c) are wrong.

Finally, as stated earlier, such methods of controlling for confounders require both prior knowledge of which variable might act as a confounder and a reliable measurement of this variable. It is almost always

difficult, if not impossible, to ever know all the variables that might affect a certain outcome. For example, in this theoretical scenario treatment duration might also be affected by several factors like the number/type of extracted teeth, the patient's compliance with aligners, or the clinician's expertise. Furthermore, treatment duration might be affected by several factors unknown to us, like genetic predisposition of a patient. Logically, one cannot 'control' statistically for a factor that is unknown. Herein lies one of the biggest advantages of randomized clinical trials – namely, the random allocation of patients into groups, which ensures that the compared groups will have a similar distribution of all known and unknown confounders at baseline. Since the current study was a retrospective non-randomized one however, statement (d) is wrong.

## References

- Greenland S and Morgenstern H (2001) Confounding in health research. *Annu Rev Public Health* 22: 189–212.
- Richmond S, Shaw WC, O'Brien KD, Buchanan IB, et al. (1992) The development of the PAR Index (Peer Assessment Rating): reliability and validity. *Eur J Orthod* 14: 125–139.
- Westfall J and Yarkoni T (2016) Statistically Controlling for Confounding Constructs Is Harder than You Think. *PLoS One* 11: e0152719.

## TABLES

**Table1.** Patient characteristics and outcome data in the two groups.

Variable	Aligners	Braces	P value
Number of patients	52	54	
Age in years – mean (SD)	14.1 (1.6)	13.5 (1.6)	0.07*
Male patients – n (%)	24 (46%)	28 (52%)	0.56 <sup>†</sup>
Baseline PAR score – mean (SD)	29.4 (11.5)	33.1 (9.4)	0.08*
Tx duration in months– mean (SD)	19.2 (5.3)	19.2 (3.6)	0.99*

PAR, Peer Assessment Rating; SD, standard deviation; Tx, treatment.

\* from t-test for independent samples.

<sup>†</sup> from chi-square test.

**Table 2.** Linear regression of the effect of appliance type (independent variable) on treatment duration in months (dependent variable).

Model	Category	$\beta$ (95% CI)	P value	Adjusted for
1	Aligners	Reference		
	Braces	-0.01 (-1.76, 1.74)	0.99	-
2	Aligners	Reference		
	Braces	0.05 (-1.73, 1.84)	0.95	Age
3	Aligners	Reference		
	Braces	0.08 (-1.66, 1.81)	0.93	Sex
4	Aligners	Reference		
	Braces	-1.30 (-2.31, -0.29)	0.01	Baseline PAR

$\beta$ , unstandardized regression coefficient; CI, confidence interval; PAR, Peer Assessment Rating.

## Figure legends

**Fig 1.** Scatter plot of treatment duration with baseline PAR scores, including a simple linear prediction.

